DOCUMENT RESUME

ED 430 020 TM 029 733

AUTHOR Miller-Whitehead, Marie
TITLE Tennessee TCAP Science Scale Scores: Implications for Continuous Improvement and Educational Reform or Is It Possible To Beat the Odds?
PUB DATE 1999-04-00
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Disadvantaged Youth; *Educational Change; Elementary Education; *High Achievement; Poverty; *School Districts; *Science Tests; *Scores; Test Results; Test Use
IDENTIFIERS Reform Efforts; *Tennessee Comprehensive Assessment Program

ABSTRACT

Evidence provided by analysis of science scale scores on the McGraw-Hill CTB/4 science test for grades 2 through 8 in Tennessee, part of the Tennessee Comprehensive Assessment Program (TCAP), shows that it is possible for high achieving school systems to show continuous improvement from year to year. These results would tend to offset fears that regression to the mean precludes the highest achieving school systems from maintaining gains over a period of several years. Results show that it is possible for schools with a high percentage of disadvantaged students also to be high achieving, although the lowest achieving school systems had the highest percentages of students on free and reduced lunch status. Results also show that over the 8-year period from 1990 to 1997 the mean science scale score by year statewide showed gains for 5 of the 8 years, with a slight decrease in 1997 prior to replacement of the CTBS/4 with the newer McGraw Hill TerraNova Test. (Contains three tables and seven references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 020

Tennessee TCAP Science Scale Scores:

Implications for Continuous Improvement and Educational Reform

or

Is It Possible to Beat the Odds?

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Marie Miller-Whitehead
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Marie Miller-Whitehead, Ph.D.

Education Consultant

Marie Miller-Whitehead
201 South Reynolds, L 105
Alexandria, VA 22304
mmmiller@quality.org or TNMarie@aol.com
703.751.0909

TM029733

Paper presented at the annual meeting of the American Educational Research Association,
Montreal, Canada, April 22, 1999

BEST COPY AVAILABLE

Abstract

Evidence provided by analysis of science scale scores on the McGraw Hill CTB/4 science test for grades 2 through 8 in Tennessee shows that it is possible for high achieving school systems to show continuous improvement from year to year. These results would tend to offset fears that regression to the mean precludes the highest achieving school systems from maintaining gains over a period of several years. Results showed that it is possible for schools with a high percentage of disadvantaged students also to be high achieving, although the lowest achieving school systems had the highest percentages of students on free and reduced lunch status.

Results also showed that over the 8 year period from 1990 to 1997 the mean science scale score by year statewide showed gains for 5 of the 8 years, with a slight decrease in 1997 prior to replacement of the CTBS/4 with the newer McGraw Hill TerraNova test.

As school systems across the United States continually seek ways to both improve and measure student achievement and to form stronger linkages between instructional delivery, student expectations, and accountability, there is a continuing need to provide parents, students, school policymakers, and the public with answers to the question, "How are our students doing?"

Unfortunately, there is more than one answer to such a question and those who seek simple answers or solutions to complex questions are doomed to disappointment. School systems and educational providers have at hand and may use criterion referenced tests, norm referenced tests, or performance assessments for a variety of purposes ranging from promotion-retention decisions and program admission, to planning for and providing services for identified at-need student populations (i.e., funding decisions).

There continues to be wide-spread debate over the use, and what many consider to be the misuse, of criterion, performance, and norm referenced tests with many school systems using a combination of two or all three methods as one component of their decisionmaking process. Such is the case in Tennessee, where state legislation mandates that schools and school systems show acceptable gains on the norm referenced core subject area subtests. Until 1998 when it was replaced by a newer test, Tennessee schools statewide administered the criterion and norm-referenced McGraw Hill CTB/4 tests for grades 2 through 8 in April of each year for the core subject areas of reading, language arts, math, science, and social studies, with score reports generated for both the criterion as well as the norm referenced portions of the tests. For science, the grade level tests consisted of 20 items with score reports on the norm referenced test including scale score, stanine, median national percentile rank, and total battery scores. Criterion referenced reports indicated whether the student was at mastery level, partial mastery, or no mastery on the individual domain

areas of the subtests.

The use of the scale score data in this study was based on the overall utility of scale scores in data analysis. Unlike percentiles, scale scores can be averaged (as can stanines) and provide a sensitive measure for comparisons of large groups where variance and range are large and group differences may be relatively small. While scale scores for different subject areas cannot be compared, they are an obvious choice for working with data and making comparisons of performance on one subtest. For the CTB/4, scale scores can range from 0 to 900 thus providing researchers with the ability to detect small differences between groups. Working with large data sets is somewhat problematic and raises such issues as the use of NHST (Nix & Barnette, 1998), and reporting of statistical significance and effect size (McLean & Ernest, 1998; Thompson, 1998; Daniel, 1998); nevertheless when looking at statewide indicators the use of large data sets is a necessity and those who regularly use these data are aware of the issues. Tennessee has some 139 school systems and the state report card data set includes mean scale scores for each school for each grade level for each year in the analysis. In addition, an overall system-wide mean is reported for each school by grade level. The value-added or gain score computations for each student, teacher, school, and system are performed at the Tennessee Value-Added Research Center in Knoxville for inclusion on the state report card. The value-added or gain score model was developed to control for student demographics and thus provide a method for school systems with large percentages of disadvantaged or at-risk student populations to demonstrate achievement based on past performance rather than on comparisons with systems serving completely different types of student populations.

Earlier published longitudinal analyses had revealed an overall upward trend or gain in the mean scale scores statewide (Miller-Whitehead, 1997, 1998 & 1998) although the upper bound of

scores had been highest in 1991 (807) and in 1993 (801) for groups of 8th grade students. The drop in achievement for schools at the upper level within their systems could have been the result of school redistricting, changes in grade configurations, new school openings, or other school system changes.

While the earlier longitudinal studies had focused on statewide data and trend analysis, the present analysis seeks to examine specific school systems at either end of the achievement spectrum. Many educators had expressed concerns that high achieving school systems would, in effect, fall victim to regression to the mean and not be able to show consistent gains while schools and systems on the lower levels of achievement would have more room to show improvement, thus in effect penalizing high achieving schools and systems under the value-added accountability mandate. The 5 school systems with highest overall mean science scale scores and the 5 school systems with the lowest overall mean science scale scores for the period of the study were used for comparison purposes to determine if, indeed, there was evidence of regression to the mean over the period of the study.

Each year the state of Tennessee provides a variety of data to the public including scale scores by system for each of the subject area tests and "state report card" data which includes system demographics. Table 1 shows comparisons for the 5 highest and 5 lowest achieving school systems for 1993 through 1997. Although system A did not have the highest upper bound mean scale score in the state for any of the years in the study, it nevertheless achieved not only the highest overall averages over grades 2 through 8, it also improved its overall achievement continuously each year for a five year period.

Table 1 provides evidence that concerns about regression to the mean may be overstated in

looking at system-wide achievement, although for identified subgroups of students, classes, or schools which have indeed "topped out" this would, of course, continue to be an issue. Table 1 shows that with the exception of one system (E), the top 5 systems in science score achievement also had among the lowest percentages of students on free and reduced lunch status in the state, with the top system (A) having the least number of students on free and reduced meal status (7%). In 1996, the Tennessee state average for free and reduced lunch was 39.1%, average per pupil expenditure was $4,713, and average per capita income was $13,726.

Table 1

Comparisons for High and Low Achieving Tennessee Systems

System	Mean Science Scale Scores by Year					% Free/ Reduced lunch 1995	$ Per pupil Expenditure '95-'96	$ Per Capita Income 1996
	'93	'94	'95	'96	'97			
A	749.4	751.1	754.5	756.8	760.2	7	4,600	26,458
B	747.2	744.3	749.3	751.1	752.8	19.4	5,529	16,426
C	743.6	741	745.2	745.9	745.7	18.7	6,794	18,199
D	739.5	741.3	743.9	748.2	745.8	17.6	5,457	26,458
E	732.5	737.2	746.2	748.5	746.7	38.6	5,445	17,263
V	706.6	707.3	710.2	707.8	706.1	63.7	5,418	20,372
W	696.5	707	702.2	716.9	712.3	48	3,993	15,379
X	706.6	706.2	707.1	697.8	703.1	85.3	4,877	14,743
Y	690.8	701.2	698.9	700	691.9	56	4,224	15,379
Z	693.8	689.9	692.8	691.8	688.8	81.5	4,853	14,090

Note. Demographic data is taken from the 1996 Tennessee state report card produced by the Department of Finance, Technology, and Accountability. Science scale scores analyses were conducted by author in previously cited studies.

System E's percent of students on free and reduced lunch (38.6%) placed it at approximately the midpoint (69 of 138 systems) for the state of Tennnessee (39.1% free and reduced), between the system with the lowest percent (7%) and the highest percent (85.3%). However, System E was able to show continuous improvement in science scores for 4 of the 5 years as seen in Table 1. System D also showed gains for 4 of the 5 years. Thus, it would appear that even top systems can and do continue to improve in achievement from year to year and that a high percentage of students on free

and reduced lunch does not per se sentence a school system to poor academic achievement.

For systems on the lower end of the achievement scale in science scores, all had large percentages of students on free and reduced lunch, with system X (85.3%) having the highest percentage in the state, and system Z (81.5%) having the second highest percentage of disadvantaged students. While these systems certainly had room for improvement, none of the five lowest achieving school systems showed a pattern or trend of continuous improvement over the period shown in Table 1. Of the lowest achieving systems most were very small systems of less than 1,000 students, although one of the highest achieving systems (E) had less than 1200 students.

Thus, while it would appear that even though science scale scores for the state of Tennessee have shown improvement from a mean of 721 in 1990 to a mean of 728 in 1997, systems on the lower end of the achievement scale have not progressed as much as those on the upper end.

Two of the five highest achieving school systems in Tennessee also had the highest per capita income in the state ($26,458) and one of the lowest achieving systems had a higher per capita income ($20,372) than three of the highest achieving systems, although this system also had a very high percentage of students on free and reduced lunch (63.7%). Because of the variations in per pupil expenditure among the systems, with some of the higher performing systems having lower per pupil expenditures than some of the lower performing systems and because one of the highest performing systems also had a large percentage of students on free and reduced lunch, a regression analysis was conducted with system science scale scores for 1996 as the dependent variable to determine to what extent percent of students on free and reduced lunch, per pupil expenditure, and per capita income of the systems in the analysis were related to mean science scale scores. The 1996 science scale scores and system demographic data were chosen as this was the year in which most school systems

achieved their highest science scores for the period of the study. Statewide, science scale scores were slightly lower in 1997 than in 1996, although many systems did achieve gains. Therefore, the decision was made use data for comparisons from the year 1996 in which the science scale scores were highest. Since not all systems in Tennessee serve grade levels 2 through 8[1], there were a total of 133 systems for which scale score data was calculated in the analysis for Table 2.

Table 2

Mean Science Scale Scores for Tennessee by Grade for 1990 through 1997
(N=133)

	Year							
Grade	1990	1991	1992	1993	1994	1995	1996	1997
8	767	763	768	771	765	772	774	770
7	756	753	758	755	753	764	761	758
6	742	741	734	746	735	747	745	748
5	728	727	727	727	733	728	731	733
4	707	709	719	716	716	715	718	716
3	690	689	691	686	699	691	699	699
2	666	668	667	663	675	669	676	673

Table 2 provides comparisons by grade level for science scale scores in the state of Tennessee for each year of this study. According to Table 2 the science scale scores for 1997, the last year of use of the CTB/4, were higher for each grade level from 2 through 8 than they were for the first year of the present study, 1990.

[1]There are several K-5 and K-6 systems.

Results of the regression analysis are shown in Table 3. The variables in the analysis yielded an R^2 of .91 and an adjusted R^2 of .86. Therefore the model was quite effective in identifying those variables which were closely related to science scale score achievement in the year 1996.

Table 3

Summary of Regression Analysis[b] for 1996 Mean Science Scale Score, Percent Free and Reduced Lunch, Per Capital Income, and Per Pupil Expenditure (N=10)

Source	SS	df	MS	F	Sig of F	R^2	Adj. R^2
Regression	5452.78	3	1817.59	19.01	.002[a]	.91	.86
Residual	573.6	6	95.6				
Total	6026.38						

[a] Predictors: Constant, Per Pupil, Per Capita, Free and Reduced
[b] Dependent Variable: Mean Science Scale Score, 1996

Of the variables in the analysis, percent free and reduced lunch was by far the most powerful predictor for school system science achievement in 1996 ($r = -.94$, $p < .001$) with per capita income of the county in which the school system was located having a positive correlation ($r = .63$, $p < .05$) to system-wide student performance on the CTB/4 and per pupil expenditure ($r = .46$, $p < .1$) also having a positive correlation to student achievement in science. Interestingly enough, there was no correlation between the per capita income of the counties in the analysis and their school system's per pupil expenditure for education. The state of Tennessee has, since the implementation of the Education Improvement Act of 1991, funded schools on the basis of fiscal capacity of the county in which they are located; one reporting area for school systems on the state report card is "effort to capacity." Thus, some counties fund their schools above 100% and some below 100%. This accounts for some wealthy systems which have lower per pupil expenditures than their poorer

neighbors. Therefore, it is not uncommon for less wealthy counties to not only receive a higher level of state funding due to poverty in the county and but also to provide more local effort than that required by the state.

Even though one system with a high percentage of disadvantaged students (eligible for the free and reduced lunch program) was able to "beat the odds," the most significant finding was the negative correlation of -.94 for percent of students on free and reduced lunch and student achievement on the CTB/4 science test. While per pupil expenditure was able to offset this disadvantage to a certain extent, nevertheless it is an uphill battle and school systems which serve at risk populations continue to be challenged by the negative effects of poverty on student achievement.

Opportunities for improvement certainly exist in respect to students and school systems on the lower end of the scale and while Tennessee's efforts to provide a more equitable education for these students are to be praised, much remains to be done. However, it is also interesting to note that on the other end of the achievement scale, systems which were doing well continued to improve. Although two of the systems in this study serve wealthy counties (as indicated by per capita income figures) not all of the high performing systems were located in the wealthiest counties. One of the 5 highest achieving systems in the state ranked 21st in the state in per capita income and the other ranked 27th, an indication that county wealth alone does not necessarily correspond to or predetermine student achievement.

References

Daniel, L. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the Schools, 5*(2), 23-32.

McLean, J. and Ernest, J. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

Miller-Whitehead, M. (1997). An analysis of science scale scores for grades 2 - 8 in Tennessee for 1990 - 1994. (ERIC Document Reproduction Service No. ED 414 160)

Miller-Whitehead, M. (1998). A longitudinal analysis of science scale scores grades 2 - 8 in Tennessee for 1992 - 1996. (ERIC Document Reproduction Service No. ED 414 170)

Miller-Whitehead, M. (1998). Tennessee TCAP science scale scores 1990 - 1997: Implications for continuous improvement and educational reform. (ERIC Document Reproduction Service No. ED 416 070)

Nix, T. and Barnette, J. (1998). The data analysis dilemma: Ban or abandon. A Review of null hypothesis significance testing. *Research in the Schools, 5*(2), 3-14.

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*(2), 33-38.